Frito-Lay Optimization Project

Prepared For: Professor Peng Shi
Prepared by: Good fun! Consulting Group

Team 10

Sharon Jia | Eva Tang | James Lee

Liana Liang | Sylar Liu | Penny Tsai

April 2019

Good fun!
consulting

Table of Contents

Table of Contents 1

Executive Summary 2

Opportunity for Improvement and justification 3

Optimization Methodology 4

I. Into Summary 4

II. Methodology 5

Explanation of Methodology 7

I.Reasoning and choice 7

II.Main assumptions: 9

III.Potential alternatives: 10

IV.Weakness of the analysis: 10

Optimization Result 11

I.Discussion of output 11

II.Clarity of results 11

Rigour of Results 12

Recommendation 12

I.Recommendation to Frito-Lay and explanation 12

II.Issues that are not captured by the model 13

Technical Appendix 14

A1. Mathematical Formulation 14

A2. Discussion of Technical Details: 15

Executive Summary

Frito-Lay, a subsidiary of PepsiCo, manufactures, distributes and sells some of America's favorite snack foods, including *CHEETOS*®, *FRITOS*®,*DORITOS*® , and *LAY'S*®. Frito-Lay has the world's largest portfolio of billion dollars food and beverages brand. Within their US market, there are over 1400 SKUs in total and 10 sales regions. Since small format accounts for 40% of business and Frito-Lay views it as their competitive edge, we designed a prototype optimization tool specifically for small format to consolidate the SKUs for each region. In the report, we only use the southwest data as sample. Because number of distinct SKUs to keep in distribution center is limited to 250 for eaches module, our optimization tool will return 250 SKUs that consist of products with high growth potential in addition to Frito-Lay's popular products.

Even though the current process of SKU selection is able to fulfill demand in each sales region across US, it is highly manual and uses a spreadsheet-based tool to rank SKUs by velocity and staleness. The current model neglect product profit margin, transportation cost and macro trend in market. When we conducted exploratory data analysis (EDA) on current 250 SKUs in southwest region's data, we caught some outliers whose return rate is relatively high but sales volume is low. The current method also ignores transportation cost because it does not optimize on route. It would not make sense to ship a product from a plant across the country when potential substitutions are available at near plants. Another drawback of current tool is that it solely depends on velocity and staleness. The growth rate of current 250 SKUs is -17%.

Based on the current model, our prototype optimization tool improved it from profit margin, transportation cost and product growth rate perspective. It maximizes profit using both profit margin and transportation cost. At the same time, it takes product growth rate into consideration by adding constraints to the objective function and using BCG matrix as SKUs filter. The final 250 SKUs improves total profit by 227%. Profit jumps from $67,002,758 to $148,576,633.

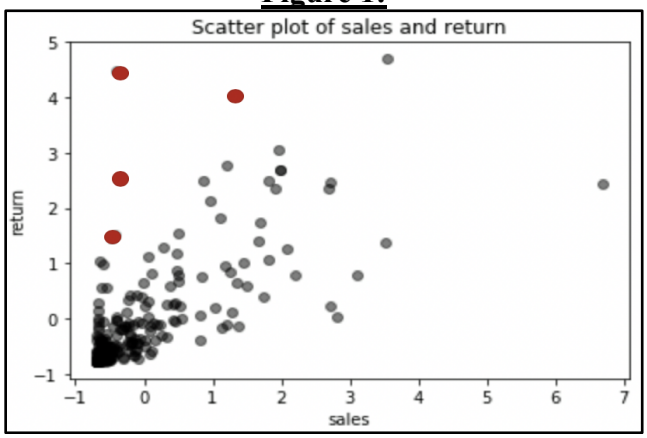
Our recommendation to the business is thoroughly discussed in the end of this report. We made our recommendations from three perspectives, data collection, inventory management and macro economic trend. In term of data collection, it is in favor of data analysts they will be working with in the future as well as the company in a sense that they can collaborate with their business partners. In term of inventory management, we suggested Frito-Lay to replace current manual process with an automated optimization tool to boost up profit in a sense of SKU selection and minimize labor cost. In term of macro economic trend, as Frito-Lay is also adopting many mega trends by changing their packaging and acquiring healthy brand, we made our optimization tool feasible for changes as long as the company run it regularly and consistently on a monthly or quarterly basis.

This report is well structured in 6 sections plus 2 technical appendix–opportunity for improvement and justification, optimization methodology, explanation of methodology, optimization result, rigour of results and recommendation. Under each section, the problems are well-defined and explained in detail with appropriate visualization. The 2 technical appendix contains the information of mathematical formulation of our optimization model and its technical details.

Opportunity for Improvement and justification

In 2018, the small format returns in southwest amount to $ 10,406,110. As the current SKUs selection is manually decided by regional manager based on velocity and stateness performance, we believe this leads to many opportunities for improvement using a data-driven optimization tools. When diving deep into the sales and returns data for our current SKUs, we discover that there are products that are clearly overly supplied, which result in rather high returns. This could be result a from the overestimated supply amount or a wrong choice of SKUs, and in both cases would indicate the opportunity for improvement. Below in there is a graph that illustrates the current inefficiencies SKU selection:

Figure 1:



The plot above shows that on the y-axis are the returns standardized and the x-axis are the sales standardized. As we see in the graph, we have marked products in red as the ones with high return and low sales. We see this as an opportunity for improvement and decided that we a part of our optimization tool that it can exclude products like these from our SKUs.

The current SKUs selection method also overlook several factors that are important to the business. First of all, it does not consider product profit margin even though profitability is essential. Secondly, the transportation could be inefficient. Alternatively, we can optimize the shipment of products by choosing efficient route and deciding whether each product should be ship to and from to avoid accruing additional cost, which is also a key metric in the business. In terms of strategic planning, the current SKUs selections are difficult to incorporate business strategy as well as consider the factor for growth in the current selection method. In fact, when we calculate the growth of revenue for each product, we find that for the current portfolio of 250 SKUs, the average growth rate is -17%, which also proves that it will be difficult to align with Frito-Lay's strategy for growth. All of these aspect shows that it will be beneficial for Frito-Lays to reform their SKUs selection by adopting a data-driven optimization tool that can optimize selection decision based on profitability while taking all the essential factors into consideration.

We used the BCG matrix on the status quo SKUs selections and got a result for the distribution in the matrix below:

For original SKU portfolio:

Dogs	6.76%
Stars	63.06%
Cash cows	25.23%
Question marks	4.95%

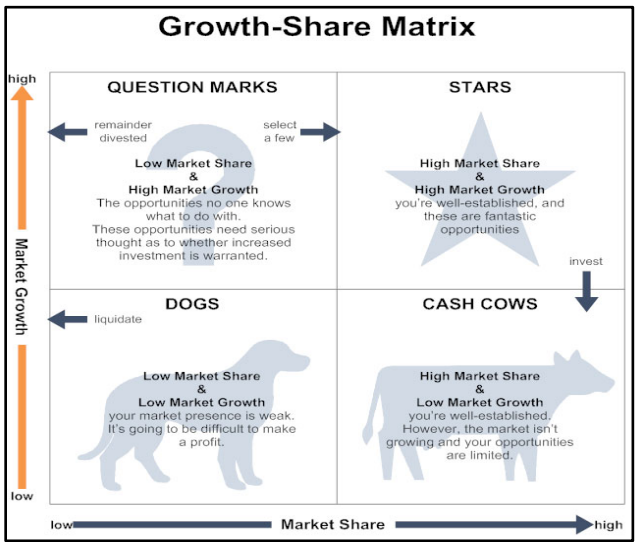
From this table we can see that there are dogs in the SKUs and is a big issue in that these should be eliminated. Also, we see that there is a large amount of the distribution in cash cows and stars while the question marks are only about 5% of the portfolio. The question marks are what become the stars and growth opportunity. By seeing this distribution, we can tell that this isn't aligned with what type of business strategy Frito-Lay wanted and would give us an opportunity to make sure our tool can accomplish this.

Optimization Methodology

I. Into Summary

From the start of our optimization, we wanted to understand what would be the best tool that we can provide to Frito-Lay. We wanted something that made business sense that would be easy to understand while having a technical model that is built using data to drive the decisions. With this all in mind, we have researched ways that would be able to accomplish both tasks that were ambitious and concluded that the BCG matrix would be the best to combine both and be the core concept for our optimization tool. With the company selling multiple products this brings a challenge of measuring the performance of each item and have difficulty keeping bias mindset out. To understand our model, we must first explain what the BCG matrix is. The BCG matrix or also known as the growth-share matrix provides a framework for analyzing products and portfolio of products according to growth and market-share being on each axis. The matrix is to intend to help with long-term strategic planning, to help companies recognize opportunities for growth through examining their product range to determine where to invest, discontinue, or develop products.

First, we collect the data on the market share which is determined by the quantities sold and the growth rate for all the products that Frito-Lay offers. The relative market share is intend to help with long-term strategic planning, to help companies recognize opportunities for growth through examining their product range to determine where to invest, discontinue or develop products.



In the graph shows multiple locations represented with a symbol that represents a category and what characteristics they have. The bottom left corner represents the dogs that have low market share and growth, question marks are the low market share with high market growth, stars are high market share with high growth, and the cash cows are high market share with low market growth. This strategic planning is regularly used for conglomerates to identify their companies however have been recently used to identify products. The life cycle of the products moves through from dogs to question marks to starts to cash cows._For the market share, we divide the total sales for each product by the total sales for all product; and for the market growth, we used the exponential growth rate for the product sales over periods except the first period as the first period data is not representative. After collecting the data on all of the products we were able to categorize the products into one of the BCG matrix.

This optimization tool aims to maximize total profit while taking market demand, distribution cost, macro trend and Frito-Lay's development strategy all into consideration. Through the data-driven optimization process, the optimized selection of 250 SKUs is provided as well as the optimized shipping quantity. Also, we made sure to consider all the important factors.

In the optimization, we first apply BCG matrix to select candidate products based on their performance on market share and growth rate. Out of the candidate products, we try to optimize the product selection as well as the shipping quantity to achieve maximizing the total profit of chosen products and minimizing its distribution cost. At the same time, we set constraints to ensure the new SKU coordinate with the structure of market demand with diversification. To incorporate Frito-Lay's development strategies in optimization, we also allow company to decide the desired proportion for cash cows, stars, question marks, and macro trend products based on the company's business strategy.

II. Methodology

Required input:

1. **Sales data:** sales information including sales_period, region, chain_id, BDC, BDC_desciption, sales, and returns. Format as "sales2018.csv".

2. **Sales product:** sales product information including BDC, BDC_description category_description, business_unit_description. Format as "sales2018_products.csv".
3. **Product data:** product information including BDC and the price retailer charged for the consumer without any promotion.
4. **Chain Information:** chain_id and the sales channel it belongs to.
5. **Distribution cost:** the cost information for each BDC shipped from one warehouse to another including the cost based on trucking distnce, cost of sending and receiving the shipment by the source and destination warehouse.
6. **Estimated Margins:** the profit margin for each BDC

Decision:

1. Whether to select certain BDC as our SKUs.
2. The amount of quantity each SKU needs to be shipped from certain plants to certain distribution centers.

Output Data: 250 selected SKUs for small format.

Objective: Maximize total profit.

Constraints:

1. For core products that brought a certain percentage of cumulative sales in the past would need to be selected in the new SKU and their quantity of supply should be greater than or equal to their previous quantity sold and less than previous quantity sold times its growth rate.
2. The quantities in each category for new SKU should be greater than the past, while less than the past quantities time the growth rate of each category.
3. The proportion for cash cows, stars, and question marks products for new SKU should equal to the ratio decided by the company.
4. The amount of health and wellness products for new SKU are selected by ranking sales, here we choose top 10 to be included in but it can be changed.
5. The total selected SKUs should be greater than or equal to 200 and less than or equal to 250.
6. The amount of quantity each BDC needs to be shipped from certain plants to certain distribution centers needs to be greater than or equal to 0.
7. For each distribution center and each BDC, the total amount received from all the plants should be the same.
8. For each plant that doesn't produce certain BDC, the amount of that BDC shipped from that plant to all distribution center should be 0.
9. The amount of quantity each SKU needs to be shipped from certain plants to certain distribution centers is an integer.
10. The decision whether to select certain BDC as our SKUs is either yes or no.

This methodology is the exact solution to what Frito-Lay is wanting in that we are considering optimizing on profit because we will be able to lower their cost and consider long-term growth to capture the most crucial part of their issue they have presented us.

Explanation of Methodology

I.Reasoning and choice

Reasoning and choice of input data:

By combining sales data, sales product, product data, and chain information, we will be able to know the demand structure for the snack market in both quantity and monetary amount which helps us to further identify the core products in small format, quantities in each category for the snack market, as well as the performance of growth for each category. With the distribution cost data, we can calculate the total cost for supplying each BDC and optimization the shipping arrangement accordingly.

Reasoning and choice of objective and decision variables:

The goal for this project is to optimize the SKU selection to avoid managerial overhead and inventory efficiency. Considering that profitability is crucial to every company, our objective is set to maximize total profit margin for each product times the quantity minus the distribution cost. Therefore, our decisions include whether to choose certain SKU and the optimized amount for shipping to achieve optimized outcome.

Reasoning and choice of constraints:

1. BCG proportion:

A constraint we included was in regard to the BCG matrix. So far, the matrix was used to select products candidates and how we want to make sure that the matrix is in the model. So, the best approach was to label each product as a category of (dogs, question marks, stars, or cash cows). With these labels, we are able to identify how many of each category is in the final SKUs. Therefore, we created a constraint that would be able to control the quantities of SKUs from each category. This means that Frito-Lay would have the option to adjust each category depending on the business strategy they would want to pursue. So, for this constraint, we can make sure that out of the SKUs that there is less than or equal to zero dogs apart of the model and that there will be more than or equal to 40% question marks and similar with cash cows and stars.

Each quantity given to each category represents a different strategy. For the BCG Matrix, there are different strategies already created. And in this step is where the business ties in with the model and data. There are four well-known strategies for the matrix:

1. Build strategy: the concept is to create a new image for the brand by targeting a new audience with products that can reach a new audience. This means that they should invest heavily in promoting a product to increase the market share. There is an opportunity for a strong campaign to moving question marks into a star's quadrant.
2. Hold strategy: for this strategy, it is to maintain the benefits from the current products and leave each product in each quadrant.
3. Divest strategy: this strategy is to abandon investment in the product that are in the dogs/failing category. The market is flooded or even the product is of little to no interest.
4. Harvest strategy: This is where you stop all investments and gain the most profit from the current cash cows.

Our strategy that is in line with Frito-Lay is to use a hybrid approach of first divesting the dogs in the current product line from the candidates of SKUs and to use a distribution of reinvesting in

question marks and stars. This would be using all strategies of build, hold, and divest as seen in figure 2.

Figure 2:

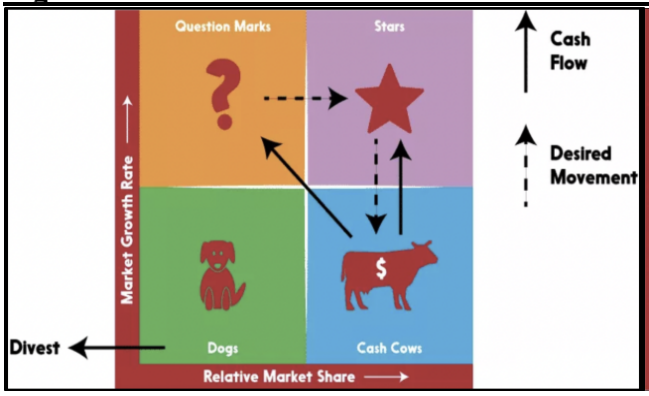


Figure 3:

This hybrid strategy as seen in figure 2 would ensure that we are taking the risks necessary in investing in question marks and convert into stars while holding cash cows for hedging against the risks from the building strategy.

For the distribution of all the four categories in our model was decided on the business strategy of PepsiCo. This decision would still be adjustable for the management team however for our recommendation we conducted research and concluded that the below graph is PepsiCo's general strategy. In Figure 3 we can see that this PepsiCo strongly values question marks and stars and makes sure to have the cash cows to back the investments. So we decided to mimic the strategy of divesting the dogs and having greater than or equal to 30% of question marks, greater than or equal to 35% of stars and greater than or equal to 35% of cash cows.

2. category quantity:

The reasoning behind setting a quantity requirement for product categories is that first of all, we believe the demand structure for different types of snack is rather stable. For example, the demand for "oat" might not be substitutable by "potato chips" unless there are dramatic changes in the market. In addition, by setting the quantities for each product category, we also ensure diversification in our new SKUs supply and that different customer segment can still be satisfied in the optimization process. When setting the quantities for each product category, it is worth noticing that we are not just using the current demand but rather, we also take the growth for each category into consideration so that the demand will be reasonable.

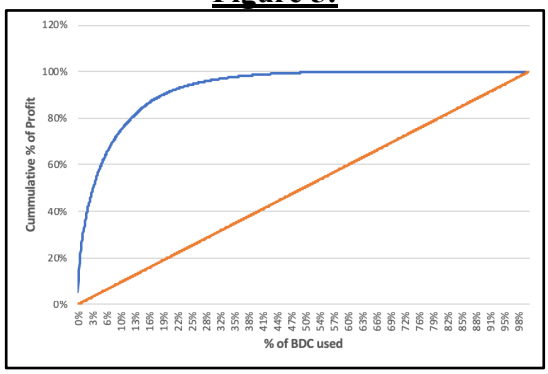
3. core products:

For one of our constraints we came up was that we want to take risks to strive for new opportunities in the dynamic market, however, be able to mitigate that risk by not losing the core products and our core business. We realize that this can be achieved through identifying the core products/business and putting a constraint on making sure those products are in the future offers. This was done by focusing our data on small retail which includes c-stores as well as small groceries because we wanted to know the core products for the SKU products. We aggregated the sales for these retailers and ranked them by the sales for each product. This would give us the core products shown in a example as figure 4. When we graph top products ranked and the cumulative sum of the percentage we get the graph in figure 5 below.

Figure 4:

BDC_description_x	revenue	cum_sum	cum_sum_percent	product_count	product_percentage
DORITOS XXL NAC	230695777	230695777	0.053573396	1	0.001015228
VPK SUPER WHTCRN	204956243	435652020	0.101169421	2	0.002030457
VPK SUPER 24CTVPK	189813052	625465072	0.145248814	3	0.003045685
LAY'S XXL REG	142846026	768311098	0.178421275	4	0.004060914
DORITOS SUPER NAC	84727621.2	853038720	0.198097173	5	0.005076142
DORITOS XXL CRA	80954616.7	933993336	0.216896883	6	0.006091371
DORITOS XXL SPY NAC	73308060.8	1007301397	0.233920869	7	0.007106599
LAY'S XXL WAVY	70563650.2	1077865047	0.250307534	8	0.008121827
CT CRY XL HOT	65184292.5	1143049340	0.265444976	9	0.009137056
MULTI SUPER FLMHOT	56788800.6	1199838140	0.278632772	10	0.010152284

Figure 5:



With this information, we can see the top 88 products bring in 75%, of sales from all small retailers. Our team identified these products as the core products and wanted to make sure though this constraint to include the 88 products by setting the constraint that the amount of new supply needs to be greater or equal to old demand to secure the market. This will overall help keep our core customers for small retailers satisfied while mitigating the risks.

4. Macro trend:

Health and wellness products are the new standard for the snack market globally, with a projected annual growth rate of 3% relative to the American snacks being around 1%. This would be a three times growth rate relative to competitors. As a market leader for snack, Frito-Lay has already created or acquired wellness brands to capture the market opportunity and we think it will be important to incorporate this strategic decision to ensure long-term profitability.

5. Limitation for SKUs:

The SKUs each warehouse can store is limited so we set the upper bound as 250 SKUs. To include flexibility while maintaining diversity, we set limitation for SKUs as a range from 200-250.

II. Main assumptions:

1. Market share for each product categories should be similar in the future.

When we selected SKU, we deliberately set the quantities of product category greater than before. The rationale behind SKU selection is that we replaced some SKUs with better substitutions. So, the substitutions should incrementally improve total profit while maintaining the original product quantities in each category.

2. The transportation cost can be inferred using business unit code.

The BCG matrix selected a total of 842 SKUs but not all of them had cost data. To fill in missing values, we assumed that all BDCs were shipped only from plants to distribution centers and

estimated their total shipment cost by averaging cost of other product with the same business unit code.

3. Future annual growth rate remains same as previous years.
Since it is hard to gauge future growth rate, we used past growth rate to project future sales with the belief that economic growth rate will not fluctuate dramatically in the near future.

4. The amount of each product distributed to each distribution center are the same.
Regardless the plants from which the distribution center received a certain BDC, every product was distributed equally across distribution centers. Due to a lack of information on annual demand, we made this assumption to reduce computational complexity. Frito-Lay can always feed the actual data in to our model to get a more precise result.

5. The 1019 SKUs in southwest region are all SKUs available for the region.
The final 250 SKUs were picked out of these 1019 SKUs in southwest region. The new 250 SKUs are 47.2% overlapped with original SKUs in southwest region, and out of the 250 SKUs, 132 are different.

6. The profit margin provided in current data has only excluded production cost.
We subtracted total revenue by total transportation cost to calculate total profit. This equation holds only when this assumption holds. Otherwise, we would exclude transportation cost twice. But can be adjusted by taking out transportation cost if so.

III. Potential alternatives:

1. Use time-series to project future sales
Time series forecast could be a potential alternatives. It can be used to predict the next period demand for each product. When we change the SKU portfolio, the demand will not stay the same. Therefore, even when we use the time series to forecast the demand, the prediction is not reliable. Hence, we used the annualized growth rate as it is less biased.

2. Incorporate actual data instead of simulation or estimation and more data.
Due to a lack of many informational data, we had to estimate the missing data such as transportation cost. For example, if demand/supply and store-level data are available, we will be able to improve accuracy substantially and present additional insights as mentioned in assumption 4 under session main assumption.

IV. Weakness of the analysis:

1. Optimization tool has to be ran continuously and on a regularly bases
Since we fed our model with current sales data to get a more decisive result, it should to be ran on a monthly or quarterly basis to get the same accurate result. In addition, since our model consider profit along with market trend, it is better to update the model regularly to keep up with macro forces. This aspect is furtherly demonstrated in recommendation section.

2. Analysis might not be robust as a result of missing data

After we merged 3 tables, 23 SKUs sales data are missing. Due to a lack of information, we cannot judge their importance. Having 23 SKUs missing can compromise our model. Please refer to technical notes for detailed explanation. We categorized retailers intuitively based on their descriptions. For example, if it is a small groceries or convenience stores, we categorized them into small formats. Since it is a subjective process, Frito-Lay might have categorized retailers differently than us; and therefore, the accuracy of our model will change once Frito-Lay feeds our model with their data and categories.

Optimization Result

I. Discussion of output

The output consists of 250 SKUs selected by the optimization model. The 250 SKUs are tailored to small formats in southwest region only, yet the optimization can be applied across all regions. As demonstrated previously, the 250 SKUs should maximize profit: total revenue minus total transportation cost, while still taking market growth into account. To simulate future total revenue, we used estimated profit margin multiplied by estimated demand. From input data listed above, we were able to calculate total transportation cost incurred.

II. Clarity of results

After running our model, we were able to get $\$1.5107 \times 10^8$ total revenue from the new 250 products. The new 250 SKUs are listed in the table below in Figure 8. This table contains all the BDC for 250 SKUs after we apply the optimization. In comparison, as you can see from the Figure 6, the previous total profit was $67,002,758 and the optimized total profit is $148,576,633, total revenue minus transportation cost, which is around 227% increase in the total profit margin. We chose to exclude the transportation cost for this number because we do not have the new transportation cost and would be an inaccurate representation.

Figure 6:

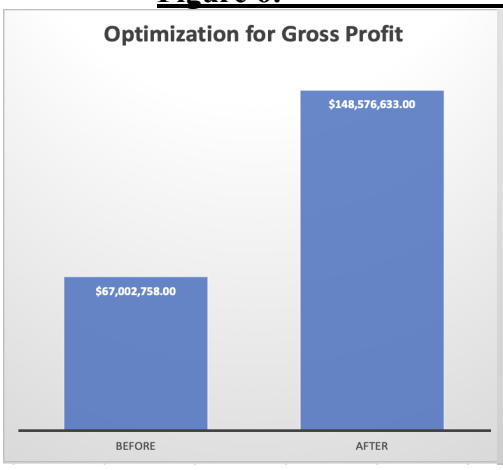


Figure 7:

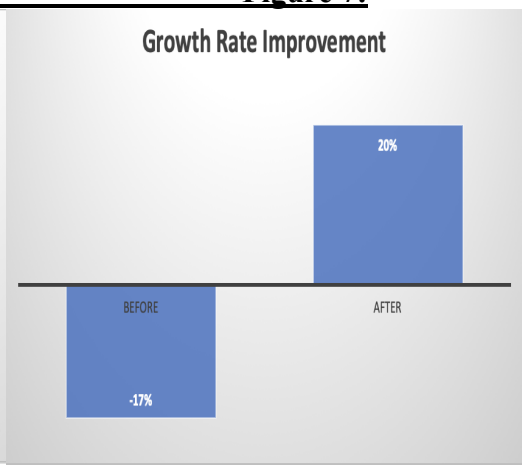


Figure 8:

4192011	90321082	2015071	60038758	13005031	49026005	20016019	15016429	2015011
4192051	90321313	2015082	60054008	13005071	49026025	20016037	15016819	2015014
28015052	90363093	2015546	60054064	13010031	49026118	20016083	15016901	90055031
28015548	45192130	2015772	60054085	13026031	61003031	20016097	15021008	90055313
700306410	45192787	2015797	60054758	13026041	61026031	20016270	15021033	90058024
3015009	705321319	2016009	172015071	13026071	61026275	20016973	15021060	90058053
3015071	23192019	2016011	1026006	13026758	63026050	20021083	28199816	15016055
3015915	23192044	2016014	1026071	14026191	105199883	20021097	33005045	15016060
3021009	114130011	2016041	1026182	14026559	126389143	35329028	40005011	15016096
13015031	114130132	2016082	2005009	14026577	126389715	35329038	40005462	28026052
13015041	114505062	2016715	2005011	14148577	61015031	20015019	40026011	28026548
13015067	85004019	2021009	2026006	15003033	61015139	20015534	45003130	28148548
13015071	85021016	2021011	2026009	15026008	61015186	20015569	12003045	6026030
13015420	85021063	2021014	2026011	15026019	61015889	45199130	12005045	6026284
13015758	85021171	2022746	2026014	15026033	61021031	48026031	12005602	9005071
13021071	85021299	12016011	2026039	15026060	81094506	49003005	12026011	5015025
13311031	85021473	12016041	2026041	15026730	48015011	12026264	12026044	5015053
105192883	85021494	12016044	2026071	15026830	48015017	12026602	12026045	5015061
105192884	85021548	12016045	2026082	15148730	48015031	12026852	5015086	6015008
105192899	85021904	12016087	2026546	20026019	48015638	5192086	5015529	6015284
43005082	85166071	12016548	2026727	20148019	48015711	60038008	5015548	7140835
1015007	85166480	12016852	2026798	22199011	48015745	60038085	5015898	10045065
1015011	85465071	12016932	2026892	23005019	14015181	2015038	5022389	90047031
1015097	85465509	12021011	3005009	23026019	15015144	2015039	5022763	90047253
1015182	85465511	12021045	3005577	23199044	15015786	2015041	7523006	90047523
1021097	85521099	49015005	3026009	24026034	15016008	90092031	7523063	90054157
8006039	106540031	49015156	3026463	28003052	15016019	90092300	9192071	
10044068	106540033	49015788	4026051	28005052	15016033	90321031	2015009	

Rigour of Results

By selecting the optimal 250 SKUs and efficient transportation routine, we maximized the difference between total gross margin and transportation cost. Compared with the previous SKUs, our optimized selection chose 134 different SKUs. With the selected 250 products, we increase the total gross margin from $67,002,758 to $148,576,633, with 227% improvement. As those 250 products are selected according to the methodology below, we make sure the products have both high profit margin and potential growth.

As you can see in the Figure 7, the previous SKUs portfolio has average growth rate of -17%, and our optimized portfolio has average growth rate of 20%. This huge improvement in the growth rate prove our hypothesis that there are indeed some opportunities for Frito-Lay to improve. Compared with the BCG matrix, the original SKUs have: qm ratio for 4.95%, cc ratio for 25.23%, star ratio for 63.06%, and dog ratio for 6.76%; the optimized SKUs have: qm ratio for 19.20%, cc ratio for 15.63%, star ratio for 65.20%, and dog ratio for 0.00%. The optimization selected more question mark products, which have high potential to grow and .The previous SKUs selection is both inefficient in the profit margin and also lack of potential growth opportunity. Our optimization help Frito-Lay identify the optimal product with both high margin and growth.

Recommendation

I. Recommendation to Frito-Lay and explanation

1. Data collecting

We suggest Frito-Lay to collect more data in favor of data analysts they will be working with in the future. For example, as we mentioned, our model will perform even better if we can feed it with actual data instead of estimations and simulations. Moreover, collecting data is favorable for the company itself in term of cooperating with their partners. In other words, if sales data is given at store level, for instance Ralph's, we are able to get the exact demand for Grandma's chocolate

chip cookie at a particular Ralph's store and use it to predict future demand for this specific product. At the same time, Frito-Lay can sell it to their partner, for example a company who produce milk, and collaborate in an ethical way. Because cookie and milk are considered compliments. They can project the demand for cook in a neighborhood using demand of cookie given by Frito-Lay's store-level data.

2. Inventory management

Although current manual process and a spreadsheet-based tool are able to fulfill demand for small formats which count as 40% of the business. After we conducted an exploratory data analysis (EDA), we detected some SKUs with high return rate but low in sales. We recommend Frito-Lay to manage their inventory automatically using our optimization tool because it will filter out SKUs with high return and low sales. Moreover, it will potentially reduce cost in a sense that manual select is costly.

3. Keep track of future trends

Health and wellness is a macro trend and Frito-Lay is also adopting it in many ways. For instance, they acquired many healthy brands such as Bare and Naked in recent years. Thus, we ranked all products under health and wellness category by their sales amount and selected the top 10.

II. Issues that are not captured by the model

Most of the following issues we encountered are due to a lack of data and trade secret in some situations. As we recommended in former part, if Frito-Lay can collect more data and feed our model actual data other than estimations, they should have a more precise result.

1. Transportation cost

Information on transportation cost are only provided from plants to distribution centers. We did not have access to transportation cost from distribution centers to retails.

2. Profit margin for every product

We only have estimated profit margin for categories of products. Under each category, their profit margin

3. Data over past few years

The current dataset is limited to 1 year. It is hard to see the trend in demand and growth without observation over time.

4. Store-level data

We are not able to include store-level data in our model because it was not provided. However, we do recommend, as mentioned above, Frito-Lay to include store-level data. It benefits the company in a way of getting better optimization results as well as cooperation with business partners.

5. Regional flavor preference

The model can work for every region, rather than for the nation, meaning it does not consider the difference in flavor preference between regions. The regional director may hope to explore the distribution of flavors preference within one region but the model cannot achieve it without detailed sales data.

6. Cannibalism

Another issue our model did not capture is that all these products in our portfolio can possible be competing with one another. This implies product cannibalism. We have tried to avoid it with the given information by diversifying our products by categories. However, with better detail and data it would be possible to analyze this in greater depth.

Technical Appendix

A1. Mathematical Formulation

Data:

Lists:

- I: the set of Plants i
- J: the set of distribution centers (DC j)
- N: the set of candidtate products identified by BDC code
- H: the set of healthy products, identified by BDC code
- $PBDC$: the set of CORE products which bring q*100% of revenue
- CC: the set of products classified as "cash cow" in BCG Matrix
- $Star$: the set of products classified as "star" in BCG Matrix
- QM: the set of products classified as "question mark" in BCG Matrix
- NC: the set of category names to group all the BDCs

Constants:

- cc: the proportion of cash cow products in SKU set, decided by Company
- st: the proportion of star products in SKU set, decided by Company
- qm: the proportion of question mark products in SKU set, decided by Company
- q: the cutoff percentage decided by Company.
- h: the number of healthy products to be included in SKU set, decided by Company
- gr: estimated growth rate on quantities of products, decided by Company

DataFrames or Dictionaries:

- M_n: the estimated margins per unit for each product n
- R_n: the conversion rates between case and stardard case for product n
- U_n: the conversion rates between unit and stardard case for product n
- $CO_{i,j}$: the cost incurred at source warehouse i for preparing the shipment + the cost incurred at the receiving warehouse j for taking the shipment and unpacking it for storage
- $CT_{i,j}$: the cost of transporting one standard case of a product from plant i to DC j.
- Q_PBDC_{pn}: the total quantities of each CORE product pn(defined in PBDC) sold in 2018
- $IsOrNot_{n,i}$: whether plant i produce product n
- $growth_{nc}$: growth rate regarding of quantities in each category nc
- GS_{nc}: groups containing BDCs that are associated with that category nc
- QC_{nc}: for each category nc, how many units of products were sold in 2018

Decision variables:

- x_n: binary variable determining whether to include product n in final SKU set.
- $a_{n,i,j}$: how many units of the product n was shipped from plant i to DC j.

Objective and constraints:

Maximize: $$\sum_{n\in N}\sum_{i\in I,j\in J} M_n a_{n,i,j} x_n - \sum_{i\in I,j\in J}\sum_{n\in N} CO_{i,j} a_{n,i,j} x_n U_n - \sum_{i\in I,j\in J}\sum_{n\in N} CT_{i,j} a_{n,i,j} R_n x_n U_n$$

subject to:

(Only 250 SKU)	$\sum_{n\in N} x_n \le 250$	for $n \in N$
(Each DC receive equal product n)	$\sum_{n\in N,j\in J} a_{n,i,j} = \sum_{n\in N,j\in J} a_{n,\hat{i},j}$	for each product n, each plant j, for each $i \in I$ and $\hat{i} \in I$ $(i \neq \hat{i})$
(Produce)	$a_{n,i,j} \le IsOrNot_{n,i} * 500000000$	for each n and plant $i \in IsOrNot$.
(Cash cow proportion)	$\sum_{n\in CC} x_n \ge (cc - 0.05) * \sum_{n\in N} x_n$	for each product $n \in CC$, for each $i \in I$ and $j \in J$.
(Star proportion)	$\sum_{n\in Star} x_n \ge (st - 0.05) * \sum_{n\in N} x_n$	for each product $n \in Star$, for each $i \in I$ and $j \in J$.
(Question mark proportion)	$\sum_{n\in QM} x_n \ge (qm - 0.05) * \sum_{n\in N} x_n$	for each product $n \in QM$, for each $i \in I$ and $j \in J$.
(Healthy products must be included)	$x_n = 1$	for $n \in H$
(Core products must be included)	$x_n = 1$	for $n \in PBDC$
(Core products grows)	$a_{n,i,j} \ge Q_P BDC_n$	for $n \in PBDC$
(Core products limits)	$a_{n,i,j} \le Q_P BDC_n * (1 + gr)$	for $n \in PBDC$
(Category sales quantities grows)	$\sum_{i\in I,j\in J} a_{n,i,j} \ge QC_{nc}$	for each category $nc \in QC$ and product $n \in NC$.
(Category sales quantities limits)	$\sum_{i\in I,j\in J} a_{n,i,j} \le QC_{nc}(1 + growth_{nc})$	for each category $nc \in QC$ and product $n \in NC$.
(Non-negativity)	$a_{n,i,j} \ge 0$	for each n, i, and j.
(Upper bound)	$a_{n,i,j} \le 5000000000$	for each n, i, and j.

A2. Discussion of Technical Details:

Lose data after merge

When analyzing original 250 SKUs, the California_eaches worksheet is inferred as it contains the present 250 SKUs, and the sales_2018 is also used to compute how much profit these SKUs created. However, these two tables have different keys: 1) product_code in California_eaches; 2) BDC in sales_2018. We use a intermediate sheet (Price_cleaned.csv) that has both 1) and 2) to merge these two tables. However, it returns that Price_cleaned doesn't include all product_code, thus leads to 23 missing BDC. This might affect the analysis for status quo.

Target on small retailers
Since the SKUs are stocked for small-format-orders, the optimization is built on the basis of the assumption: only the sales records associated with small retailers' chain id are considered. If the chain id is not representing convenience store or independent business, etc., then that record is viewed as being sold to supermarket. However, this part is not shown as a parameter, so if the company want to modify store types, they have to deep dive into function file.

Estimate quantities
The objective in the optimization is maximizing profit, which leads to an estimation on the amount of products being shipped and sold in the future. It is hard to predict future performance based on limit historical records. In this optimization, the average sales and returns in the past 4 sales period are calculated and serve as part of the estimation. We also set a parameter identifying average growth power of trends so the quantities could approximately project a future performance.

Shipment assumption
In the assumption, we assume within the same region, all distribution centers will receive same amount of each product, mainly due to lack of annual demand information. Since all the distribution centers in one region will use same 250 SKUs, it is reasonable to suppose equal demands of each distribution center for a certain product. Besides, this assumption simplifies optimization process and significantly saves time.

Flexible cutoff
This optimization tool sets multiple unfixed parameters that company would like to change based on domain knowledge. However, aiming to present a quantitative result, these numbers are settled depending on our exploration of data. If change the parameters, the results might be different.